

A Appendices

A.1 Hyper-parameters and Training Details

In our experiments, the **BiLSTM+MLP** predicate tagging model only takes words and lemmas as input, and its encoder structure is the same as our main model, so the hyper-parameters are also consistent with our main model. With the **BERT+MLP** predicate tagging model, the motivation for choosing this instead of using BERT as embedding in the **BiLSTM+MLP** architecture is to achieve fair comparability with the results of (Zhou et al., 2019).

For the hyper-parameters of our main model, we borrowed most parameter settings from (Dozat and Manning, 2017; Wang et al., 2019), including dropout and initialization strategies. Hyper-parameters for our baseline and proposed high-order model are shown in Table 5. We use 100-dimensional Glove (Pennington et al., 2014) pre-trained word embeddings for English and 300-dimensional FastText embeddings (Bojanowski et al., 2017; Grave et al., 2018) for all other languages. As for the pre-training, ELMo (Peters et al., 2018) is only used in English, we take the weighted sum of the 3 layers as the final features, while different versions of BERT (Devlin et al., 2019) are used in different languages, as shown in Table 6, we always use the second-to-last layer outputs as the pre-trained features.

Following the work of (Wang et al., 2019), during model training, the training strategy includes two phases. In the first phase, we used Adam (Kingma and Ba, 2014) and annealed the learning rate 0.5 every 10,000 steps. When the training reaches 5,000 steps without improvement, the model optimization enters the second phase; the Adam optimizer is replaced by AMSGrad (Reddi et al., 2018). We trained the model for maximum 100K update steps with batch sizes of {4K, 2K, 3K, 4K, 6K, 6K} tokens for CA, CS, DE, EN, ES, JA, and ZH, respectively. The training is terminated with an early stopping mechanism when there is no improvement after 10,000 steps on the development sets.

A.2 Detail Results

The proverb that *there is no such thing as a free lunch* tells us that no method works in every condition and scope. We explore our proposed high order structure learning for SRL in different languages and conditions: using pre-training or not, given or

Hidden Layer	Hidden Sizes
Word Embedding	100 (en) / 300 (others)
Lemma Embedding	100
Predicate Indicator / Sense Emb	50 / 50
ELMo/BERT Linear	100
Stacked BiLSTM	3 × 600
Biaffine Arc/Label Scorer	600
Triaffine Arc Scorer	150
Dropouts	Dropout Prob.
Word/Lemma/Predicate	20%
BiLSTM (FF/recur)	45%/25%
Biaffine Arc/Label Scorer	25%/33%
Triaffine Arc Scorer	25%
Optimizer & Loss	Value
Balance param λ	0.1
Adam β_1	0
Adam β_2	0.95
Learning rate	$1e^{-2}$
LR decay	0.5
L2 regularization	$3e^{-9}$

Table 5: Hyper-parameters for baseline and high-order SRL models in our experiment.

	Version	Provider
CA	<i>multi_cased.L-12_H-768_A-12</i>	(Devlin et al., 2019)
CS	<i>Slavic BERT: slavic_cased.L-12_H-768_A-12</i>	(Arkhipov et al., 2019)
DE	<i>multi_cased.L-12_H-768_A-12</i>	(Devlin et al., 2019)
EN	<i>www_uncased.L-24_H-1024_A-16</i>	(Devlin et al., 2019)
ES	<i>BETO: spanish.www_cased.L-12_H-768_A-12</i>	(Cañete et al., 2020)
JA	<i>NICT BERT: japanese_100k.L-12_H-768_A-12</i>	(NICT, 2020)
ZH	<i>chinese.L-12_H-768_A-12</i>	(Devlin et al., 2019)

Table 6: BERT versions for different languages.

not given predicates, and different high-order structure combinations. We report all possible results on development sets, in-domain test sets, and out-of-domain test sets in Tables 7, 8, 9 and 10. The experimental results illustrate the following points:

1. In different languages, combinations of high-order structures bring different improvements. Some high-order structure combinations are even worse for performance in some languages.

2. Pre-training can bring about a significant improvement in performance on both in-domain and out-of-domain test sets; however, the in-domain improvement is significantly greater than that of out-of-domain when the two domains are far apart. In particular, the difference between in-domain and out-of-domain in German and English is large, while the two domains in Czech are similar.

3. The SRL results in German are lower than in other languages, the data analysis found that the proportion of predicates is very small, resulting in the sparse targets, which can not train the model well, especially when no predicates are pre-identified.

Language	Method	Dev			Test			OOD		
		P	R	F ₁	P	R	F ₁	P	R	F ₁
CA	baseline	85.62	83.66	84.63	85.85	84.09	84.96			
	+sib	85.91	83.98	84.94	86.09	84.46	85.27			
	+cop	85.78	83.83	84.79	85.97	84.17	85.06			
	+gp	86.24	83.25	84.72	86.11	83.51	84.79			
	+sib+cop	85.70	84.25	84.97	85.90	84.85	85.37			
	+sib+gp	85.71	84.00	84.85	85.81	84.19	84.99			
	+cop+gp	86.05	83.98	85.00	85.74	83.72	84.72			
	+ALL	85.64	84.09	84.86	85.67	84.24	84.95			
	baseline ^{+B}	86.95	85.74	86.34	87.10	85.72	86.40			
	+sib	86.94	85.92	86.43	87.15	86.05	86.59			
	+cop	87.15	85.96	86.55	87.14	86.00	86.57			
	+gp	87.17	86.00	86.58	86.87	85.93	86.40			
	+sib+cop	86.96	86.57	86.76	86.89	86.53	86.71			
	+sib+gp	87.26	86.55	86.90	87.10	86.40	86.75			
	+cop+gp	87.24	86.12	86.68	87.08	85.97	86.52			
	+ALL	87.49	86.11	86.79	87.52	86.29	86.90			
CS	baseline	91.20	89.79	90.49	90.87	89.49	90.18	91.22	89.88	90.54
	+sib	91.30	89.71	90.50	91.08	89.56	90.32	91.25	89.72	90.48
	+cop	91.40	89.77	90.58	91.09	89.56	90.32	91.27	89.75	90.51
	+gp	91.35	89.63	90.48	91.16	89.40	90.27	91.31	89.62	90.46
	+sib+cop	91.36	89.88	90.81	91.33	89.89	90.60	91.50	90.03	90.75
	+sib+gp	91.26	89.68	90.47	91.09	89.48	90.28	91.29	89.79	90.53
	+cop+gp	91.24	89.61	90.42	91.01	89.51	90.26	91.23	89.73	90.48
	+ALL	91.18	89.81	90.49	90.96	89.65	90.30	91.15	89.85	90.49
	baseline ^{+B}	92.12	91.09	91.61	91.98	90.99	91.48	91.98	91.23	91.60
	+sib	92.31	91.62	91.96	91.98	91.23	91.60	91.94	91.50	91.72
	+cop	92.11	91.27	91.69	92.08	91.25	91.66	91.97	91.49	91.73
	+gp	92.02	91.19	91.60	91.97	91.20	91.58	91.85	91.32	91.59
	+sib+cop	92.04	91.30	91.67	92.38	91.49	91.93	91.84	91.47	91.65
	+sib+gp	92.11	91.24	91.68	92.06	91.15	91.61	91.84	91.41	91.63
	+cop+gp	91.99	91.32	91.65	91.94	91.25	91.60	91.87	91.61	91.74
	+ALL	92.03	91.26	91.65	91.99	91.14	91.56	91.75	91.32	91.53
DE	baseline	75.83	72.51	74.13	77.48	74.61	76.02	71.34	67.73	69.49
	+sib	76.63	73.36	74.96	77.01	75.54	76.27	71.66	69.36	70.49
	+cop	74.43	72.05	73.22	76.73	74.98	75.85	69.84	68.55	69.19
	+gp	75.69	73.02	74.33	76.33	75.11	75.71	69.74	67.35	68.53
	+sib+cop	76.24	73.25	74.72	77.53	75.33	76.41	71.28	68.88	70.06
	+sib+gp	75.29	73.02	74.14	75.86	74.74	75.29	70.15	67.79	68.95
	+cop+gp	76.22	72.05	74.08	77.00	74.25	75.60	69.99	66.76	68.33
	+ALL	75.13	72.57	73.83	76.79	74.18	75.46	71.46	67.46	69.40
	baseline ^{+B}	84.48	82.70	83.58	85.77	84.66	85.21	71.77	70.02	70.88
	+sib	83.87	83.15	83.51	84.97	85.34	85.15	72.14	71.86	72.00
	+cop	84.58	83.04	83.80	84.93	85.09	85.01	71.48	71.59	71.53
	+gp	83.82	83.15	83.49	85.01	84.53	84.77	71.80	70.67	71.23
	+sib+cop	84.67	83.61	84.14	85.82	85.27	85.54	72.23	71.48	71.85
	+sib+gp	84.46	82.58	83.51	85.21	84.10	84.65	71.93	69.04	70.45
	+cop+gp	83.49	82.30	82.89	85.20	84.41	84.80	70.88	69.69	70.28
	+ALL	84.53	82.70	83.60	84.95	84.17	84.56	71.39	68.17	69.74
EN	baseline	90.15	86.27	88.17	91.29	88.00	89.61	81.37	77.12	79.19
	+sib	89.94	86.67	88.27	91.46	88.53	89.97	81.76	77.85	79.76
	+cop	90.03	86.59	88.27	91.40	88.45	89.90	81.89	78.07	79.94
	+gp	89.86	86.13	87.96	91.41	88.14	89.75	82.20	78.51	80.31
	+sib+cop	90.20	86.97	88.55	91.33	88.74	90.02	81.80	78.36	80.04
	+sib+gp	90.00	86.89	88.42	91.21	88.58	89.87	81.38	77.90	79.60
	+cop+gp	89.69	86.58	88.11	91.26	88.68	89.95	81.32	78.02	79.64
	+ALL	90.03	86.91	88.44	91.60	88.95	90.26	82.6	78.75	80.63
	baseline ^{+B}	91.35	88.84	90.08	92.31	90.18	91.23	86.14	83.49	84.79
	+sib	91.3	89.11	90.19	92.49	90.58	91.53	85.96	83.97	84.95
	+cop	91.3	89.19	90.23	92.21	90.5	91.35	86.03	84.07	85.04
	+gp	91.66	88.6	90.11	92.77	90.09	91.41	86.22	83.1	84.63
	+sib+cop	91.16	89.6	90.37	92.36	90.84	91.6	85.59	83.92	84.75
	+sib+gp	91.62	88.99	90.28	92.64	90.44	91.53	86.33	83.61	84.95
	+cop+gp	91.27	89.02	90.14	92.37	90.35	91.35	86.03	83.88	84.94
	+ALL	91.56	89.35	90.44	92.59	90.98	91.77	86.49	83.80	85.13

Table 7: w/ pre-identified predicate results.

Language	Method	Dev			Test			OOD		
		P	R	F ₁	P	R	F ₁	P	R	F ₁
ES	baseline	84.58	82.58	83.57	84.97	82.60	83.77			
	+sib	84.90	83.11	83.99	85.18	83.28	84.22			
	+cop	84.66	82.76	83.70	85.12	83.21	84.15			
	+gp	84.95	82.33	83.62	85.51	82.31	83.88			
	+sib+cop	84.66	82.98	83.81	85.36	83.45	84.39			
	+sib+gp	84.73	83.13	83.92	85.05	83.21	84.12			
	+cop+gp	84.80	82.42	83.59	85.35	82.87	84.09			
	+ALL	84.89	82.90	83.89	85.12	83.29	84.20			
	baseline ^{+B}	87.14	85.91	86.52	87.23	85.98	86.60			
	+sib	87.36	85.62	86.48	87.48	85.97	86.72			
	+cop	87.03	85.94	86.48	87.19	86.11	86.65			
	+gp	87.21	86.04	86.62	87.22	85.95	86.58			
	+sib+cop	86.98	86.45	86.71	87.24	86.67	86.96			
	+sib+gp	87.62	85.66	86.63	87.62	85.84	86.72			
	+cop+gp	87.26	85.85	86.55	87.09	85.82	86.45			
	+ALL	87.49	86.11	86.79	87.52	86.29	86.90			
JA	baseline	88.49	76.68	82.16	88.15	77.79	82.65			
	+sib	87.30	78.45	82.64	86.14	79.85	82.88			
	+cop	87.71	77.22	82.13	87.90	78.58	82.98			
	+gp	86.65	77.15	81.63	86.03	78.62	82.16			
	+sib+cop	87.97	78.78	83.12	87.51	79.38	83.25			
	+sib+gp	88.32	77.72	82.68	88.34	78.51	83.14			
	+cop+gp	88.36	76.99	82.28	88.09	78.04	82.76			
	+ALL	88.86	77.18	82.61	88.17	78.28	82.93			
	baseline ^{+B}	89.93	80.89	85.17	89.63	81.83	85.55			
	+sib	89.29	81.06	84.98	89.02	82.08	85.41			
	+cop	89.99	80.56	85.02	89.71	81.57	85.45			
	+gp	89.43	80.40	84.67	88.88	81.08	84.80			
	+sib+cop	88.67	82.34	85.39	88.65	83.32	85.90			
	+sib+gp	89.90	80.05	84.69	89.75	81.27	85.30			
	+cop+gp	89.20	81.27	85.05	88.59	82.03	85.19			
	+ALL	90.51	80.33	85.12	89.69	81.66	85.49			
ZH	baseline	87.28	83.84	85.52	87.95	83.63	85.73			
	+sib	87.58	83.96	85.73	87.94	83.80	85.82			
	+cop	88.33	83.40	85.80	88.61	83.29	85.87			
	+gp	88.08	82.25	85.07	88.46	82.01	85.12			
	+sib+cop	87.95	83.81	85.83	88.09	83.64	85.81			
	+sib+gp	88.28	83.42	85.78	88.20	83.26	85.66			
	+cop+gp	88.38	82.56	85.37	88.54	82.63	85.48			
	+ALL	88.44	83.40	85.85	88.35	83.82	86.02			
	baseline ^{+B}	89.63	86.69	88.13	89.94	86.60	88.24			
	+sib	89.47	87.40	88.42	89.64	87.34	88.48			
	+cop	89.63	87.35	88.48	89.79	87.33	88.54			
	+gp	89.16	86.74	87.93	89.54	86.69	88.09			
	+sib+cop	89.80	86.92	88.34	89.97	87.45	88.69			
	+sib+gp	89.60	87.39	88.48	89.79	87.30	88.53			
	+cop+gp	89.48	87.18	88.32	89.70	87.36	88.52			
	+ALL	88.95	87.58	88.26	89.07	87.71	88.38			

Table 8: w/ pre-identified predicate results.

Language	Method	Dev			Test			OOD		
		P	R	F ₁	P	R	F ₁	P	R	F ₁
CA	baseline	83.97	82.62	83.29	84.45	82.93	83.69			
	+sib	84.17	82.97	83.57	84.57	83.21	83.89			
	+cop	84.21	82.85	83.52	84.38	82.94	83.66			
	+gp	84.52	82.23	83.36	84.55	82.28	83.40			
	+sib+cop	84.03	83.15	83.59	84.69	83.46	84.07			
	+sib+gp	84.15	83.06	83.60	84.35	83.07	83.70			
	+cop+gp	84.44	82.93	83.68	84.32	82.64	83.47			
	+ALL	83.95	83.07	83.51	84.14	83.10	83.62			
	baseline ^{+B}	85.05	84.47	84.76	85.51	84.73	85.12			
	+sib	85.14	84.66	84.90	85.60	85.10	85.35			
	+cop	85.46	84.80	85.13	85.57	84.99	85.28			
	+gp	85.33	84.81	85.07	85.29	84.93	85.11			
	+sib+cop	85.39	85.27	85.33	85.47	85.38	85.42			
	+sib+gp	85.39	85.27	85.33	85.47	85.38	85.42			
	+cop+gp	85.39	84.89	85.14	85.53	85.05	85.29			
	+ALL	86.08	85.35	85.72	86.15	85.49	85.82			
CS	baseline	90.25	88.84	89.54	89.98	88.47	89.22	89.98	88.47	89.22
	+sib	90.37	88.78	89.57	90.17	88.52	89.34	89.89	88.44	89.16
	+cop	90.43	88.82	89.62	90.16	88.53	89.34	89.86	88.45	89.15
	+gp	90.38	88.70	89.53	90.21	88.34	89.26	89.86	88.30	89.07
	+sib+cop	90.73	88.94	89.82	90.21	88.65	89.42	89.85	88.49	89.16
	+sib+gp	90.34	88.77	89.55	90.15	88.41	89.27	89.88	88.51	89.19
	+cop+gp	90.32	88.72	89.51	90.12	88.47	89.29	89.86	88.44	89.14
	+ALL	90.28	88.90	89.58	90.25	88.68	89.45	89.82	88.85	89.33
	baseline ^{+B}	91.32	90.42	90.87	91.25	90.20	90.72	90.97	90.30	90.63
	+sib	91.29	90.61	90.95	91.22	90.45	90.83	90.89	90.56	90.72
	+cop	91.33	90.59	90.96	91.32	90.50	90.91	90.99	90.58	90.78
	+gp	91.22	90.52	90.87	91.21	90.45	90.83	90.87	90.37	90.62
	+sib+cop	91.25	90.63	90.94	91.23	91.21	91.22	90.84	90.55	90.69
	+sib+gp	91.31	90.55	90.93	91.31	90.37	90.84	90.82	90.48	90.65
	+cop+gp	91.18	90.63	90.91	91.20	90.47	90.83	90.84	90.66	90.75
	+ALL	91.22	90.55	90.89	91.23	90.37	90.80	90.74	90.42	90.58
DE	baseline	53.59	68.87	60.27	51.03	72.95	60.06	39.97	45.14	42.40
	+sib	53.81	69.89	60.81	51.32	73.63	60.48	40.26	46.01	42.94
	+cop	52.34	68.70	59.41	50.41	72.70	59.54	39.82	45.79	42.60
	+gp	53.42	69.72	60.49	50.72	73.38	59.98	40.19	45.30	42.59
	+sib+cop	53.28	70.29	60.61	50.79	73.44	60.05	40.67	45.95	43.15
	+sib+gp	53.15	69.72	60.32	50.21	72.58	59.36	39.92	45.19	42.39
	+cop+gp	53.13	68.53	59.86	51.17	72.77	60.09	40.14	44.87	42.37
	+ALL	53.10	69.21	60.09	50.67	72.46	59.63	40.59	44.98	42.67
	baseline ^{+B}	57.87	80.14	67.21	55.67	83.18	66.70	37.99	43.13	40.40
	+sib	57.68	80.76	67.30	55.70	83.98	66.98	39.00	44.38	41.51
	+cop	57.71	80.71	67.30	55.66	83.61	66.83	38.00	44.05	40.81
	+gp	57.52	80.54	67.11	55.35	83.12	66.45	38.53	43.89	41.04
	+sib+cop	58.00	81.33	67.71	55.86	84.17	67.15	38.94	44.38	41.48
	+sib+gp	58.18	80.36	67.50	55.43	82.44	66.29	38.66	43.07	40.75
	+cop+gp	57.23	79.97	66.71	55.43	82.99	66.47	38.34	42.86	40.47
	+ALL	58.38	80.48	67.67	55.57	82.62	66.45	37.67	42.15	39.78
EN	baseline	85.18	82.58	83.86	86.12	85.34	85.73	74.51	73.48	73.99
	+sib	85.36	83.21	84.27	86.00	85.64	85.82	74.38	73.31	73.84
	+cop	85.31	82.91	84.09	86.12	85.56	85.84	74.35	73.07	73.70
	+gp	85.40	82.67	84.01	86.04	85.13	85.58	74.43	72.68	73.55
	+sib+cop	85.15	83.17	84.15	86.26	86.06	86.16	74.76	73.65	74.20
	+sib+gp	85.16	83.22	84.18	85.82	85.59	85.71	74.18	72.92	73.55
	+cop+gp	84.93	83.07	83.99	86.00	85.59	85.79	74.29	73.31	73.80
	+ALL	85.10	83.00	84.04	86.16	85.56	85.86	74.65	73.17	73.90
	baseline ^{+B}	88.21	85.64	86.90	88.51	88.05	88.28	80.49	79.65	80.07
	+sib	88.19	85.88	87.02	88.66	88.39	88.52	80.41	79.84	80.13
	+cop	88.13	86.00	87.05	88.39	88.30	88.34	80.32	80.26	80.29
	+gp	88.54	85.41	86.95	89.01	87.98	88.49	80.63	79.24	79.93
	+sib+cop	88.01	86.40	87.20	88.55	88.60	88.57	79.87	79.89	79.88
	+sib+gp	88.38	85.78	87.06	88.82	88.19	88.50	80.57	79.55	80.06
	+cop+gp	88.16	85.79	86.96	88.59	88.20	88.40	80.35	80.11	80.23
	+ALL	87.98	86.25	87.11	88.77	88.62	88.70	80.01	79.80	79.90

Table 9: w/o pre-identified predicate results.

Language	Method	Dev			Test			OOD		
		P	R	F ₁	P	R	F ₁	P	R	F ₁
ES	baseline	83.35	81.79	82.57	83.52	81.58	82.54			
	+sib	83.72	82.42	83.06	83.66	82.26	82.96			
	+cop	83.59	82.09	82.84	83.70	82.15	82.91			
	+gp	83.74	81.57	82.64	83.98	81.18	82.55			
	+sib+cop	83.41	82.11	82.75	83.87	82.38	83.11			
	+sib+gp	83.56	82.35	82.95	83.49	82.18	82.83			
	+cop+gp	83.65	81.61	82.62	83.87	81.86	82.85			
	+ALL	83.72	82.09	82.90	83.64	82.18	82.90			
	baseline ^{+B}	85.88	85.21	85.55	85.92	85.09	85.50			
	+sib	85.99	84.89	85.44	86.19	85.13	85.66			
	+cop	85.81	85.28	85.55	85.90	85.27	85.58			
	+gp	85.85	85.30	85.57	85.84	85.12	85.48			
	+sib+cop	85.66	85.64	85.65	86.21	85.75	86.00			
	+sib+gp	86.20	84.93	85.56	86.32	85.02	85.66			
	+cop+gp	85.88	85.09	85.48	85.76	84.94	85.35			
	+ALL	86.08	85.35	85.72	86.15	85.49	85.82			
JA	baseline	79.86	68.68	73.85	79.67	68.53	73.68			
	+sib	78.92	70.12	74.26	77.77	70.01	73.69			
	+cop	79.51	68.92	73.83	79.48	69.23	74.00			
	+gp	79.23	69.39	73.98	78.14	69.24	73.42			
	+sib+cop	79.78	70.47	74.84	79.14	69.57	74.05			
	+sib+gp	79.99	69.22	74.22	79.77	68.98	73.98			
	+cop+gp	80.45	69.04	74.31	79.91	69.01	74.06			
	+ALL	80.37	69.06	74.28	80.04	69.16	74.20			
	baseline ^{+B}	81.20	74.84	77.89	82.03	74.24	77.94			
	+sib	82.77	74.49	78.41	80.53	76.86	78.65			
	+cop	82.61	74.35	78.27	82.33	75.36	78.69			
	+gp	82.15	74.16	77.95	80.98	75.48	78.13			
	+sib+cop	81.55	75.84	78.59	82.16	75.85	78.88			
	+sib+gp	82.28	73.62	77.71	82.10	75.04	78.41			
	+cop+gp	82.14	74.89	78.35	81.35	75.73	78.44			
	+ALL	82.72	74.21	78.23	81.88	75.39	78.50			
ZH	baseline	82.00	80.20	81.09	83.08	79.91	81.46			
	+sib	82.34	80.42	81.37	83.04	80.00	81.50			
	+cop	83.04	79.81	81.40	83.65	79.50	81.52			
	+gp	82.84	78.73	80.73	83.71	78.25	80.89			
	+sib+cop	82.70	80.19	81.43	83.07	80.99	82.01			
	+sib+gp	82.87	79.89	81.35	83.31	79.52	81.37			
	+cop+gp	82.97	79.07	80.98	83.63	78.87	81.18			
	+ALL	83.19	79.86	81.49	83.51	79.59	81.50			
	baseline ^{+B}	86.42	84.20	85.29	86.82	83.98	85.38			
	+sib	86.24	84.78	85.51	86.47	84.63	85.54			
	+cop	86.34	84.72	85.52	86.55	84.54	85.53			
	+gp	85.90	84.24	85.07	86.37	84.00	85.17			
	+sib+cop	86.53	84.61	85.56	86.83	84.57	85.68			
	+sib+gp	86.33	84.74	85.53	86.64	84.58	85.60			
	+cop+gp	86.28	84.63	85.45	86.61	84.66	85.63			
	+ALL	85.63	84.99	85.31	85.90	84.90	85.39			

Table 10: w/o pre-identified predicate results.